



DR WPI: 96-342288/34.  
 DR N-PSDB: T34137.  
 PT Production of icosapentaenoic acid using transformed E. coli - uses  
 PT DNA coding for icosapentaenoic acid synthase derived from Shewanella  
 PT strain  
 PS Claim 7: Page 128-131; 145pp; English.  
 CC The DNA sequence (T34137) which encodes the biosynthetic enzymes of  
 CC icosapentaenoic acid (EPA) can be used to transform Escherichia coli.  
 CC The DNA sequence allows efficient microbial production of EPA, which  
 CC is a raw material for drugs, agrochemicals, foods and animal  
 CC feedstuffs. EPA is also useful for lipid balance correction and as  
 CC an antihypertensive, antiinflammatory and anticancer agent.  
 SQ Sequence 543 AA;  
 SQ 69 A; 28 R; 25 N; 26 D; 0 B; 4 C; 24 Q; 33 E; 0 Z; 41 G; 9 H;  
 SQ 27 I; 46 L; 25 K; 20 M; 17 F; 26 P; 30 S; 28 T; 8 W; 19 Y; 38 V;  
 Found using 'claim3' (new.key)

20 snisfdvqvmegq[kdfracyvvnhdhg  
 30 39

-----|  
 1 match found in sequence:  
 W37053 ; S. putrefaciens EPO biosynthesis gene cluster ORF9 product.  
 (from "A-Geneseg 35.2")  
 ID W37053 standard; Protein; 543 AA.  
 AC W37053;  
 DT 03-JUL-1998 (first entry)  
 DE S. putrefaciens EPO biosynthesis gene cluster ORF9 product.  
 KW SCRC-2874; FERM BP-1625; eicosapentaenoic acid; EPA;  
 KM biosynthesis gene cluster; synthetase.  
 OS Shewanella putrefaciens.  
 PN W09801565-A1.  
 PD 15-JAN-1998.  
 PF 09-JUL-1997; J02371.  
 PR 10-JUL-1996; JP-180845.  
 PA (SAGA ) SAGAMI CHEM RES CENTRE.  
 PI Kato S, Kondo K, Yamada A, Yazawa K;  
 DR WPI: 98-101060/09.  
 DR N-PSDB: V00503.  
 PT Eicosapentaenoic acid produced by culture of transformed Escherichia  
 PT coli containing an eicosapentaenoic acid synthetase gene derived  
 PT from the marine microorganism Shewanella  
 PS Example 1: Pages 93-97; 110pp; Japanese.  
 CC The present sequence is encoded by the Shewanella putrefaciens  
 CC SCRC-2874 (FERM BP-1625) eicosapentaenoic acid (EPA) biosynthesis  
 CC gene cluster.  
 CC A novel EPA (useful in drugs, pesticides, foods and feedstuffs) is  
 CC encoded by synthetase enzyme gene sequences comprising parts of the  
 CC full sequence of the synthetase gene from the marine microorganism  
 CC S. putrefaciens SCRC-2874 (FERM BP-1625), in which at least 1 of  
 CC the 9 open reading frames (ORF) (numbered 2-10) in the gene have  
 CC been deleted. In particular the gene sequences comprising the  
 CC following parts of the full gene:  
 CC (1) bases 8081-9441, 12314-13084 and 13889-32520;  
 CC (2) bases 8081-9441, 12314-13084, 13889-32520 and 34627-35559;  
 CC (3) bases 8081-9441, 12314-13084 and 13889-35559;  
 CC (4) bases 8081-9441, 9681-13084 and 13889-32520;  
 CC (5) bases 8081-9441, 9681-13084, 13889-32520 and 34627-35564; and  
 CC (6) bases 8081-9441, 9681-13084 and 13889-35564, are claimed.  
 SO Sequence 543 AA;  
 SO 69 A; 28 R; 25 N; 26 D; 0 B; 4 C; 24 Q; 33 E; 0 Z; 41 G; 9 H;  
 SO 27 I; 46 L; 25 K; 20 M; 17 F; 26 P; 30 S; 28 T; 8 W; 19 Y; 38 V;  
 Found using 'claim3' (new.key)

20 snisfdvqvmegq[kdfracyvvnhdhg

30 39  
 -- Search Statistics --  
 Times: CPU  
 00:02:10.08 Total Elapsed  
 00:02:42.00  
 Number of sequences searched: 170751  
 Number of sequence hits: 4  
 Number of separate matches: 4  
 Number of sequence hits saved: 0

> 0 <  
01 10 Intelligenetics  
> 0 <

Quest - Quick User-directed Expression Search Tool  
Release 5.4

-- Outline of search "claim3iss" --

Selected search type is key against sequence data banks or files.

Selected scope is Sequence.

Selected sequence key from "new.key":  
claim3 (AA) ID claim3 AA preliminary pattern

followed by  
1 e  
2 g or s or h or y or e  
2 f or t or s or p or n or l  
2 n or s or k or m or l or d  
2 d or k or t or e  
2 f or l or r or i  
2 t or s or n  
2 r  
2 i or v or a

Selected data banks and files:

Data bank : Issued\_AA , all entries

-- Output Parameters --

Format Options:	File Options:	No
Nucleic acid code matching	Indirect file	No
Find non-matching hits only	Sequence or key file	No
Report key used	List of hits	Yes
Note position of hit	Hit display	Yes
Display full annotations	Name and annotations	Yes
Sequence context		10

-- Run Parameters --

Run mode	Batch
Time to start comparison	now
Notify at end of run	No

1 match found in sequence:

US-08-375-709-17 ; Sequence 17, Application US/08375709  
(from "/sich/laa/5A\_COMB.pep")

Sequence 17, Application US/08375709  
Patent No. 5683898

GENERAL INFORMATION:

APPLICANT: YAZAWA, Kazunaga

APPLICANT: YAMADA, Akiko

APPLICANT: KATO, Seishi

APPLICANT: KONDO, Kiyosi

TITLE OF INVENTION: Gene Coding for Eicosapentaenoic Acid

TITLE OF INVENTION: Synthesizing Enzymes and Process for Production of

TITLE OF INVENTION: Eicosapentaenoic Acid

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/375,709

FILING DATE: 20-JAN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/178,251

FILING DATE: 14-MAY-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 4-147945

FILING DATE: 15-MAY-1992

ATTORNEY/AGENT INFORMATION:

NAME: WEGNER, Harold C.

REGISTRATION NUMBER: 25,258

REFERENCE/DOCKET NUMBER: 53466/150/AAOK

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 543 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

Found using 'claim3' (new.key)

20 SNISFDVQVMEQOLKDFSRACVYVNHADHG

30 39

1 match found in sequence:

US-08-752-929-17 ; Sequence 17, Application US/08752929  
(from "/sich/laa/5A\_COMB.pep")

Sequence 17, Application US/08752929  
Patent No. 5798259

GENERAL INFORMATION:

APPLICANT: YAZAWA, Kazunaga

APPLICANT: YAMADA, Akiko

APPLICANT: KATO, Seishi

APPLICANT: KONDO, Kiyosi

TITLE OF INVENTION: Gene Coding for Eicosapentaenoic Acid Synthesizing

TITLE OF INVENTION: Enzymes and Process for Production of Eicosapentaenoic

TITLE OF INVENTION: Acid

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/752,929

FILING DATE: 20-NOV-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/375,709

FILING DATE: 20-JAN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/178,251

FILING DATE: 14-MAY-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 4-147945

FILING DATE: 15-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: WEGNER, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 53466/150/AAOK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 543 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
Found using 'claim3' (new.key)

20 SNISFDVQVMEQOLKDFSRACYVNHADHG  
30 39

-----1  
1 match found in sequence:  
US-08-266-311-1; Sequence 1, Application US/08266311  
(from "/src/h/laa/5b.COMB.pep")  
Sequence 1, Application US/08266311  
Patent No. 5814475

GENERAL INFORMATION:  
APPLICANT: Niepel, Frank  
APPLICANT: Fleckenstein, Bernhard  
TITLE OF INVENTION: Human Herpesvirus Type 6 Protein P100,  
TITLE OF INVENTION: The Corresponding DNA Sequences, Their Preparation and Use  
NUMBER OF SEQUENCES: 13  
CURRENT APPLICATION DATA:  
ADDRESSSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSSEE: Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/266,311  
FILING DATE: 08-JUL-1991  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/908,041  
FILING DATE: 06-JUL-1992  
APPLICATION NUMBER: DE 91 111 338.9  
FILING DATE: 08-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Finnandi, Carol P.  
REGISTRATION NUMBER: 32,220  
REFERENCE/DOCKET NUMBER: 02481-1194-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 870 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
Found using 'claim3' (new.key)

110 ENMFTNKEKESQFSDINRALIRGNFIKW  
120 129

-----1  
1 match found in sequence:  
US-08-467-527A-1; Sequence 1, Application US/08467527A  
(from "/src/h/laa/5b.COMB.pep")  
Sequence 1, Application US/08467527A  
Patent No. 5827519

GENERAL INFORMATION:  
APPLICANT: Niepel, Frank  
APPLICANT: Fleckenstein, Bernhard  
TITLE OF INVENTION: Human Herpesvirus Type 6 Protein  
TITLE OF INVENTION: P100, The Corresponding DNA Sequences, Their Preparation  
NUMBER OF SEQUENCES: 13  
CURRENT APPLICATION DATA:  
ADDRESSSEE: Finnegan, Henderson, Farabow, Garrett & Dunner,  
ADDRESSSEE: Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,527A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/266,311  
FILING DATE: 27-JUN-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/126,435  
FILING DATE: 24-SEP-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/908,041  
FILING DATE: 06-JUL-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 91 111 338.9  
FILING DATE: 08-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Forman, David S.  
REGISTRATION NUMBER: 33,694  
REFERENCE/DOCKET NUMBER: 02481-1194-04000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 870 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
Found using 'claim3' (new.key)

110 ENMFTNKEKESQFSDINRALIRGNFIKW  
120 129

1 match found in sequence:  
 US-09-080-793-10 ; Sequence 10, Application US/09090793  
 (from "/srch/1aa/6A.COMB.pep")  
 Sequence 10, Application US/09090793  
 Patent No. 6140486  
 GENERAL INFORMATION:  
 APPLICANT: Calgene, LLC  
 TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression  
 TITLE OF INVENTION: of polyketide-like synthesis genes in plants  
 FILE REFERENCE: CGNE.131.0105  
 CURRENT APPLICATION NUMBER: US/09/090,793  
 EARLIER FILING DATE: 1998-06-04  
 EARLIER APPLICATION NUMBER: 60/048,650  
 EARLIER FILING DATE: 1997-06-04  
 NUMBER OF SEQ ID NOS: 66  
 SOFTWARE: Patentln Ver. 2.0  
 SEQ ID NO 10  
 LENGTH: 543  
 TYPE: PRT  
 ORGANISM: Shewanella putrefaciens  
 Found using 'claim3' (new.key)

20 SNISFDVQWMEQOLKDFSRACYVNHADHG  
 30 39

1 match found in sequence:  
 US-08-467-528-1 ; Sequence 1, Application US/08467528  
 (from "/srch/1aa/6A.COMB.pep")  
 Sequence 1, Application US/08467528  
 Patent No. 6174685

GENERAL INFORMATION:  
 APPLICANT: Neipel, Frank  
 APPLICANT: Fleckenstein, Bernhard  
 TITLE OF INVENTION: Human Herpesvirus Type 6 Protein  
 TITLE OF INVENTION: P100, The Corresponding DNA Sequences, Their  
 NUMBER OF SEQUENCES: 13  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Finnegan, Henderson, Farabow,  
 STREET: 1300 I Street, N.W., Suite 700  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20005-3315  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentln Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/467,528  
 FILING DATE: 06-JUN-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/266,311  
 FILING DATE: 27-JUN-1994  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/126,435  
 FILING DATE: 24-SEP-1993  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/908,041  
 FILING DATE: 06-JUL-1992  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: DE 91 111 338.9  
 FILING DATE: 08-JUL-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Forman, David S.  
 REGISTRATION NUMBER: 33,694  
 REFERENCE/DOCKET NUMBER: 02481-1194-03000  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-408-4000  
 TELEFAX: 202-408-4400  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 870 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 Found using 'claim3' (new.key)

110 ENMFTRKEKFEPSQSDINRLRLGNFTKW  
 120 129

1 match found in sequence:  
 US-09-231-899-10 ; Sequence 10, Application US/09231899  
 (from "/srch/1aa/6B.COMB.pep")  
 Sequence 10, Application US/09231899  
 Patent No. 6566583  
 GENERAL INFORMATION:  
 APPLICANT: Lassner, Michael  
 APPLICANT: Metz, James G  
 APPLICANT: Faciolotti, Daniel  
 TITLE OF INVENTION: SCHIZOCHTRIUM PKS GENES  
 FILE REFERENCE: CGNE.131.0205  
 CURRENT APPLICATION NUMBER: US/09/231,899  
 EARLIER FILING DATE: 1999-01-14  
 EARLIER APPLICATION NUMBER: 60/048,650  
 EARLIER FILING DATE: 1997-06-04  
 EARLIER APPLICATION NUMBER: 09/090,793  
 EARLIER FILING DATE: 1998-06-04  
 NUMBER OF SEQ ID NOS: 86  
 SOFTWARE: Patentln Ver. 2.0  
 SEQ ID NO 10  
 LENGTH: 543  
 TYPE: PRT  
 ORGANISM: Shewanella putrefaciens  
 Found using 'claim3' (new.key)

20 SNISFDVQWMEQOLKDFSRACYVNHADHG  
 30 39

-- Search Statistics --  
 Times: CPU  
 00:08:15.98  
 Total Elapsed  
 00:13:49.00  
 Number of sequences searched: 328807  
 Number of sequence hits: 7  
 Number of separate matches: 7  
 Number of sequence hits saved: 0



! FINDPATTERNS on pir:\* allowing 0 mismatches

1 E(Q,S,H,Y,E)(S,Q,I,T,N,P)(F,T,S,P,L,I)(N,S,K,M,T,D)(D,K,T,E)(F,L,R,I)(T,S,N)R

← pattern awarded (column 3)

XPBE12 ck: 1901 len: 871 ! major antigenic structural protein p100 - H

1 E(Q,S,H,Y,E)(S,Q,I,T,N,P)(F,T,S,P,L,I)(N,S,K,M,T,D)(D,K,T,E)(F,L,R

E(S)(Q)(F)(S)(D)(I)(N)R(A) pattern matched

121: NKEKF ESOFSDINRA LRLTG matching portion of db seq

D97264, ck: 8688 len: 497 ! galactose-1-phosphate uridylyltransferase [im

226: IILNN EHIPMKISRI TPENL E(Q,S,H,Y,E)(S,Q,I,T,N,P)(F,T,S,P,L,I)(N,S,K,M,T,D)(D,K,T,E)(F,L,R

T30186 ck: 3609 len: 543 ! hypothetical protein 8 - Shewanella sp.

1 E(Q,S,H,Y,E)(S,Q,I,T,N,P)(F,T,S,P,L,I)(N,S,K,M,T,D)(D,K,T,E)(F,L,R

30: DVOYM EQQLKDFSR A CYVN E(Q)(O)(L)(K)(D)(F)(S)R(A)

B90396 ck: 6793 len: 220 ! hypothetical protein SSO2259 [imported] - S

1 E(Q,S,H,Y,E)(S,Q,I,T,N,P)(F,T,S,P,L,I)(N,S,K,M,T,D)(D,K,T,E)(F,L,R

137: LINHM EELINDLSRI DLTNF E(E)(I)(T)(N)(D)(L)(S)R(I)

A89606 ck: 9797 len: 692 ! protein F18G5.4 [imported] - Caenorhabditis

1 E(Q,S,H,Y,E)(S,Q,I,T,N,P)(F,T,S,P,L,I)(N,S,K,M,T,D)(D,K,T,E)(F,L,R

653: QHEIR EELIDERSRI QCSGN E(E)(I)(L)(D)(E)(R)(S)R(I)

Databases searched: NBRF, Release 76.1, Released on 12May2003, Formatted on 10Jun2003

Total finds: 5

Total length: 96,168,682

Total sequences: 283,308

CPU time: 02:52.29





TOIG of: a89606 check: 9797 from: 1 to: 692  
PI:A89606 - protein F18G5.4 [imported] - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 10-May-2001 #sequence-revision 10-May-2001 #text-change 09-Nov-2001  
C:Accession: A89606  
R:Anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for  
investigating biology  
A:Reference number: A75000; MID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and  
www.sanger.ac.uk/Projects/C\_elegans/ for a list of authors  
A:Note: Published errata appeared in Science 283, 35, 1999; Science 283, 2103,  
1999; and Science 285, 1493, 1999  
A:Accession: A89606  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-692 <STO>  
A:Cross-references: GB:chr\_X; PIDN:AAA81080.1; PID:g1055091; GSPDB:GN00028;  
CSP:F18G5.4  
C:Genetics:  
A:Gene: F18G5.4  
A:Map position: X  
A89606 Length: 692 August 12, 2003 12:02 Type: P Check: 9797 ..  
MKHVASLHACPFHIGRCKLAPRKLLEPRHRSKLVRRALPLLSFAAFPLFCAPFHTTFAHTTY  
VAREDLILLPTICCARPHHRSSNMILKOSVOSTGPQPPKOKHNEHTTCTTKTAPSVORBS  
SAVIMPHDORLEQALIKSGISTEBALYRSLDRTVEKDVPCIHHSQPTNVTFGILLNVEMD  
ENQALTYRSMINIMMPRLSNLSWSEIKAIYIPARIRKPDIIIVNNQSTNSDACLHAYISASEP  
AAIREYASLVSTDVAVTSDGVNVLFSALFRSSCPIRVRYPPDDQCDLKFASMSHDITEINLGLMTD  
KGLDSYMNSEFEDVMTAVREVYTFPDSTNSDPIIVIRIHMHRPLFYFNHIVPCVLISMAVILGF  
LMPETGEKINIIITLLSMGYVLOSITESIPTESEGVILIGMYVSSLLAVCLATCVNITILNMRNGA  
ANOGHHVPMAMQKWTIGYATFMRSIRREDSIALIKASOSKSTIRRSILRDLKRVKNSNVRKSKKE  
QANNRCECMDLVHRIYASISCLAADTKPMSGTIRDFASESTFLGRVYSDCIMPRISSSNSVITE  
FETRRRLIKRYRSLQHEIRKEILIDERSIQCSGNNLHSLIDIFYVFPALQHCOSOSASAI

TOIG of: b90396 check: 6793 from: 1 to: 220  
PI:B90396 - hypothetical protein SS02259 [imported] - *Sulfolobus solfataricus*  
C:Species: *Sulfolobus solfataricus*  
C:Date: 24-May-2001 #sequence-revision 24-May-2001 #text-change 24-May-2001  
C:Accession: B90396  
R:She, O.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez,  
M.U.; Chan-Weiher, C.C.Y.; Clausen, I.G.; Curtis, B.A.; De Moers, A.; Erasuo,  
G.; Fletcher, C.; Gordon, P.M.K.; Heikamp-de Jong, I.; Jeffries, A.C.; Kozera,  
C.U.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.; Schenk, M.E.;  
Theriault, C.; Tolstrup, N.; Charlebois, R.L.; Doolittle, W.F.; Duguet, M.;  
Gaasterland, T.; Garrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
submitted to Genbank, April 2001  
A:Description: *Sulfolobus solfataricus* complete genome.  
A:Accession: B90396  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-220 <KUN>  
A:Cross-references: GB:AE006641; MID:g13815561; PIDN:AAK42425.1; GSPDB:GN00155  
C:Genetics:  
A:Gene: SS02259  
B90396 Length: 220 August 12, 2003 12:01 Type: P Check: 6793 ..  
MTNCEFYALDNLKDEKLSNKLIDYINTDRGLIDYIKGLIEDENTIGKIIGSLTSDDVLELLVMD  
KVIKTKLFIEMDIYNIQIFLINIDVRSKGLIDPIGLLEDEESLGIILNLINDFTNLINHEEIT  
NDLSRIDLTNKKYITLLVNSATGELAKTENVKPITSIMYIKLKLKDPIDQGLGVAASVLRKIRGLVYPOK  
GLAEVEKKLL

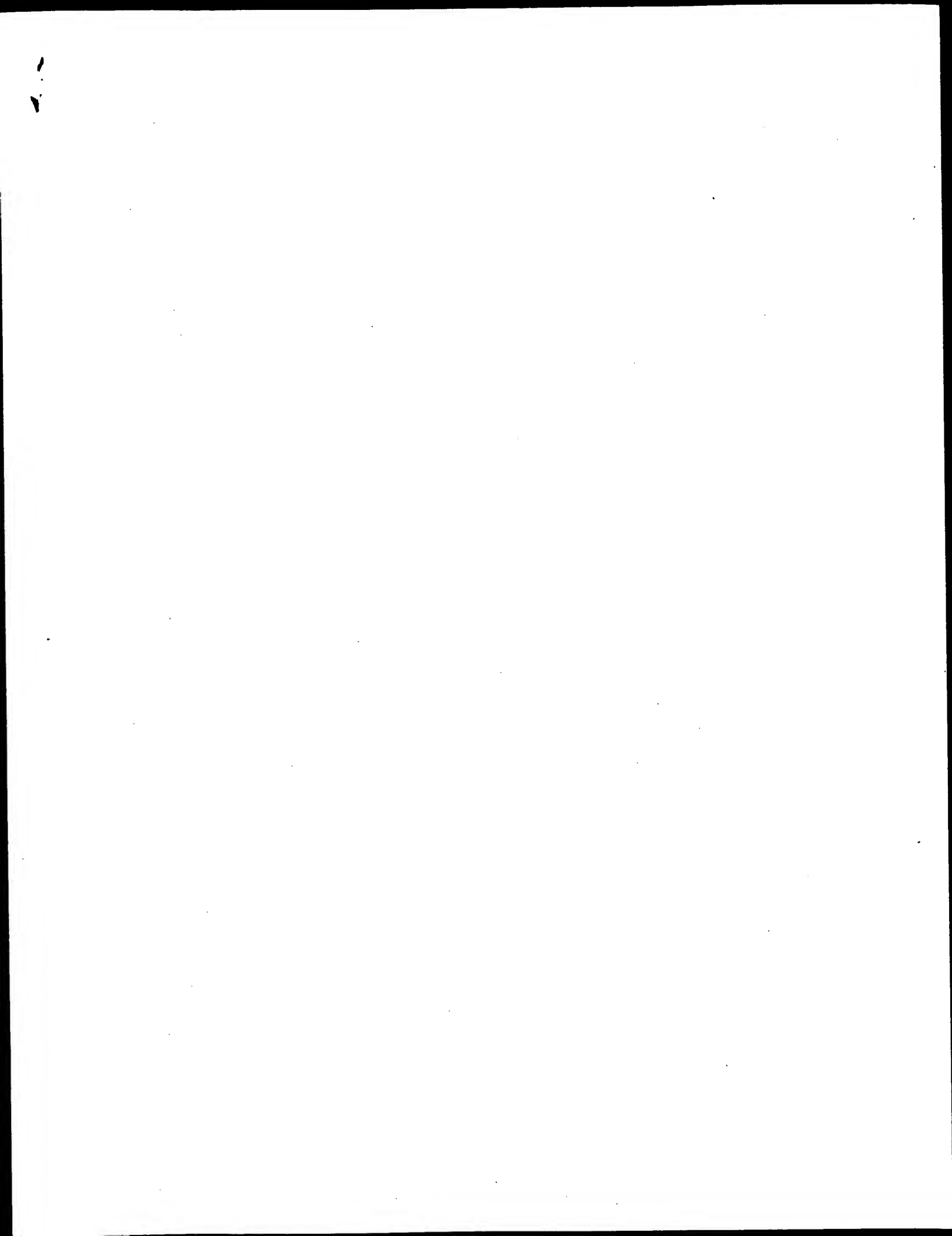
TOIG of: d97264 check: 8688 from: 1 to: 497  
P1:D97264 - galactose-1-phosphate uridylyltransferase [imported] - Clostridium  
acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: D97264  
R:Nolling, J.; Bretton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson,  
R.; Lee, H.M.; Dubois, J.; Qiu, D.; Hittl, J.; Wolf, Y.I.; Tatusov, R.L.;  
Sabathe, F.; Doucette-Stamm, L.; Soucaille, P.; Daly, M.J.; Bennett, G.N.;  
Koonin, E.V.; Smith, D.R.  
J: Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing  
Bacterium Clostridium acetobutylicum.  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: D97264  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-497 <KUR>  
A:Cross-references: GB:AE001437; PID:AAK80903.1; PID:g15026015; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics: CAC2961  
C:Superfamily: galactose-1-phosphate uridylyl transferase  
D97264 Length: 497 August 12, 2003 12:00 Type: P Check: 8688 ..  
MINHEINKLLASLKGLIQEDDKYSSNMLAGLENDNFYFEIISDVPSTATILNOLLAAYAKENLIN  
DVAEDLDPTKIMCVMPRESEVINNFENRLNNSPKKATSYKLSIASNTIRKDRIDKNTWKTPEY  
GDLDITINLSKPEKDPRIAKAKLSKSTSPKCLCKENEGFYGNINHPAROTLRIFLPLINKSKWFLQY  
SPYTYNEHCIIINNEHIMKISRITFENLISFIDILPHYFAGSNADLPIVGGSLSDHYOGGRTTAM  
EKAPVEKYSIGYEDISVGRYKMPMSYRISRSKTKLINAEHILSMRYSKDTOSLSHTSEPHN  
TTPJARKNEEYEDIVRNRRGDPNPLGIFPHNEVHHIKENIGLIEVGLAVLPARKSELALIK  
EMLIEKKIISNDSTISKNTWTKYILDNKINISENIDCIIKKEVGIRFLEVLKHAGVFRNSGLSAF  
DKFINLLI

TOIG of: t30186 check: 3609 from: 1 to: 543  
P1:T30186/- hypothetical protein 8 - Shewanella sp.  
C:Species: Shewanella sp.  
C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
C:Accession: T30186  
R:Takeyama, H.; Takeeda, D.; Yazawa, K.; Yamada, A.; Matsunaga, T.  
Microbiology 143, 2725-2731, 1997  
A:Title: Expression of the epsilon-pentanoic acid synthesis gene cluster from  
Shewanella sp. in a transgenic marine cyanobacterium, Synechococcus sp.  
A:Reference number: 220764; MUID:97419510; PMID:9274025  
A:Accession: T30186  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-543 <TAK>  
A:Cross-references: EMBL:U73935; MID:92529413; PID:92529421; PIDN:AA81126.1  
A:Experimental source: strain SCRC-2738  
T30186 Length: 543 August 12, 2003 12:00 Type: P Check: 3609 ..  
MNPATNEMLSMPMPMAVTEESNISFDVYMEQLKDESRACYVNHADHFGIAQTADIVTEQANSTDL  
VSAFTPALGTSIGDNFRFRVHGVKIVYVAGAMANGISSELYVALQAGIILGSRGAAGLIPSRVEAAI  
NRIOALDNGEYMFNLHSPSEPALERGSVELPLKHVRTVEASAFGLTPQIIVYRAAGLSRDAOGKIV  
VGNVIAKVSRTVEAKEMFAPAKMLQKLVDDGSIAPQMEALQVPMADDTAEADSGSDNRPVIT  
LLPTILALKEBIOAKYOYDPIRVGCGGVTGTPDALATFNMGAAYIVGSSINQACVENCASDHTKLLA  
TTEMADVTMAPADMFEGYKLVGVKRGTLFPMRANKLVIYTRYSIEHPIIDEREKLEKOVFRSLDE  
IMAGTVAHFNDRDPKOIERAEGNPKRKMALIFRYTIGLSRWSNSGVEGEMDYQIWAGPALGAFNOMAK  
GSYLDNODRNAVLAHKLHATGAAYLNRTNSLTAQGVKPAQLLRKPNORMAI

```

; TOIG of: xpbel2 check: 1901 from: 1 to: 871
;
; P1:XPBE12 - major antigenic structural protein p100 - human herpesvirus 6
; (strain UI102)
; C:Species: human herpesvirus 6
; C:Date: 30-Jun-1993 #sequence_revision 29-Oct-1999 #ext_change 21-Jul-2000
; C:Accession: T09303; A42533
; R:Nicholas, J.; Martin, M.
; J. Virol. 68, 597-610, 1994
; A:Title: Nucleotide sequence analysis of a 38.5-kilobase-pair region of the
; genome of human herpesvirus 6 encoding human cytomegalovirus immediate-early
; gene homologs and transactivating functions.
; A:Reference number: Z16644; MUID:94118404; PMID:8289364
; A:Accession: T09303
; A:Status: preliminary; translated from GB/EMBL/DBJ
; A:Molecule type: DNA
; A:Residues: 1-871 <NIC>
; A:Cross-references: EMBL:L25528; NID:9451932; PIDN:AAA16716.1; PID:9451934
; R:Neipel, F.; Ellinger, K.; Fleckenstein, B.
; J. Virol. 66, 3918-3924, 1992
; A:Title: Gene for the major antigenic structural protein (p100) of human
; herpesvirus 6.
; A:Reference number: A42533; MUID:92260671; PMID:1374813
; A:Accession: A42533
; A:Molecule type: DNA
; A:Residues: 2-871 <NEI>
; A:Cross-references: GB:M87287; NID:9330673; PIDN:AAA46012.1; PID:9330674
; C:Genetics:
; A:Gene: PILF1
; C:Superfamily: human herpesvirus large structural phosphoprotein; large
; structural phosphoprotein homology
; C:Keywords: phosphoprotein
; F:7-368/Domain: large structural phosphoprotein homology <CLS>
;
; XPBE12 Length: 871 August 12, 2003 12:00 Type: P Check: 1901
; xpbel2
; MMDLORHPFPAWIDRQKVEITDPLSLNLERLDNVDLREHPVNSCVREGDDVDIKTLNLLVLM
; YHYVSKRPPYNAIMODITLQSVNYEYLNKSGIKNGIFENMFNKEKESQPSDINRALLRGNFIKW
; GSNVAIDIPYVNLDAEDSEIENNLQDAEKNMWTVYNINDPMDENGYLITSINKLIYGLKFLALLOS
; WSKLEKVMASQIVITQNLHSGHRLRHDFNIVSHRYQPTLQGVESFLKITSDDYDIKSSLESLSA
; SKAFMSSEIGNSLNDIPYPLRGDIHNSNLTLPMSIDYKSSSLDPAKLKSNRSLSDSFLMQRQPKFLEL
; DSVNAGEKILKEATLGEENVKATTPASSVSLMSGVESSESTSTNDLPLSFTSTNDLRKSHGNY
; KIGPSGILDFNVKPPNANLNTNGVDLDDKSTGSSGCTDYNNGFANLNHONKSNVSPWSRNTAA
; NADFLDPYHREVPROTCPPYVNNSDVAGSEAKHTTSTETGVSFRNVFLIKDLRGKDGFRKQKSDIPK
; SLTERNDKAIMHSEVYIGDSGDATETVGARNSPALRKIKQANDFAGLNKNDRDVLRGGKNSKDLHS
; GGNKKKEMSGKFNDDKEMTNGGQEPSKSLMGDARNAGDEQYIQAGLQGVNNLLSFTNLLISGEKIE
; DILNORGTLEKLATENKSGRESEANVEKILEVSNPDQMKNFRLNDDLSVQSPRLPDADISREIDS
; ASFPDADLKLPGNGEREIDALALEKVGGETETSDLKVGDSESVPAQLKVEETPERKDDIIEOMVLRIR
; ODGETENWVSGPVAESLDIEAKGESALIASI

```



! FINDPATTERNS on swp:\* allowing 0 mismatches

1 E(Q,S,H,Y,E)(S,Q,I,T,N,P)(F,T,S,P,L,I)(N,S,K,M,T,D)(D,K,T,E)(F,L,R,I)(T,S,N,H)

GALT\_CLOAB ck: 8688 len: 497 ! Q97ez4 clostridium acetobutylicum. galactos

1 E(Q,S,H,Y,E)(S,Q,I,T,N,P)(F,T,S,P,L,I)(N,S,K,M,T,D)(D,K,T,E)(F,L,R,I)  
E(H)(I)(P)(M)(K)(I)(S)R(I)  
226: ILINN EHIPKISRI TFENL

P100\_HSV6U ck: 500 len: 870 ! Q00701 human herpesvirus (type 6 / strain u

1 E(Q,S,H,Y,E)(S,Q,I,T,N,P)(F,T,S,P,L,I)(N,S,K,M,T,D)(D,K,T,E)(F,L,R,I)  
E(S)(Q)(F)(S)(D)(I)(N)R(A)  
120: NKEKF ESQPSDINRA LRLG

PFDA\_METKA ck: 9337 len: 157 ! Q8tuy7 methanopyrus kandleri. prefoldin alfa

1 E(Q,S,H,Y,E)(S,Q,I,T,N,P)(F,T,S,P,L,I)(N,S,K,M,T,D)(D,K,T,E)(F,L,R,I)  
E(S)(S)(I)(S)(E)(L)(N)R(V)  
39: QIDLI ESSISELNRY EETLK

YZN4\_CAEEL ck: 9797 len: 692 ! P54245 caenorhabditis elegans. hypothetical

1 E(Q,S,H,Y,E)(S,Q,I,T,N,P)(F,T,S,P,L,I)(N,S,K,M,T,D)(D,K,T,E)(F,L,R,I)  
E(E)(I)(L)(D)(E)(R)(S)R(I)  
653: QHEIR EELIDERSRI QCSGN

O33907 ck: 3609 len: 543 ! O33907 shewanella sp. scrc-2738. hypothetical

1 E(Q,S,H,Y,E)(S,Q,I,T,N,P)(F,T,S,P,L,I)(N,S,K,M,T,D)(D,K,T,E)(F,L,R,I)  
E(Q)(Q)(L)(K)(D)(F)(S)R(A)  
30: DVOVM EOQLKDFSRA CYVN

Q9ESP3 ck: 4988 len: 1,851 ! Q9ESP3 rattus norvegicus (rat). mucin (frag

1 E(Q,S,H,Y,E)(S,Q,I,T,N,P)(F,T,S,P,L,I)(N,S,K,M,T,D)(D,K,T,E)(F,L,R,I)  
E(S)(N)(S)(T)(T)(L)(S)R(V)  
130: LRRGL ESNSTLSRV IMKLD

Q97WG2 ck: 6793 len: 220 ! Q97wg2 sulfolobus solfataricus. hypothetical

1 E(Q,S,H,Y,E)(S,Q,I,T,N,P)(F,T,S,P,L,I)(N,S,K,M,T,D)(D,K,T,E)(F,L,R,I)  
E(E)(I)(I)(N)(L)(S)R(I)  
137: LINHM EEINDLSRI DLTNF

Databases searched:

SWISS-PROT, Release 41.1, Released on 6Jun2003, Formatted on 9Jun2003

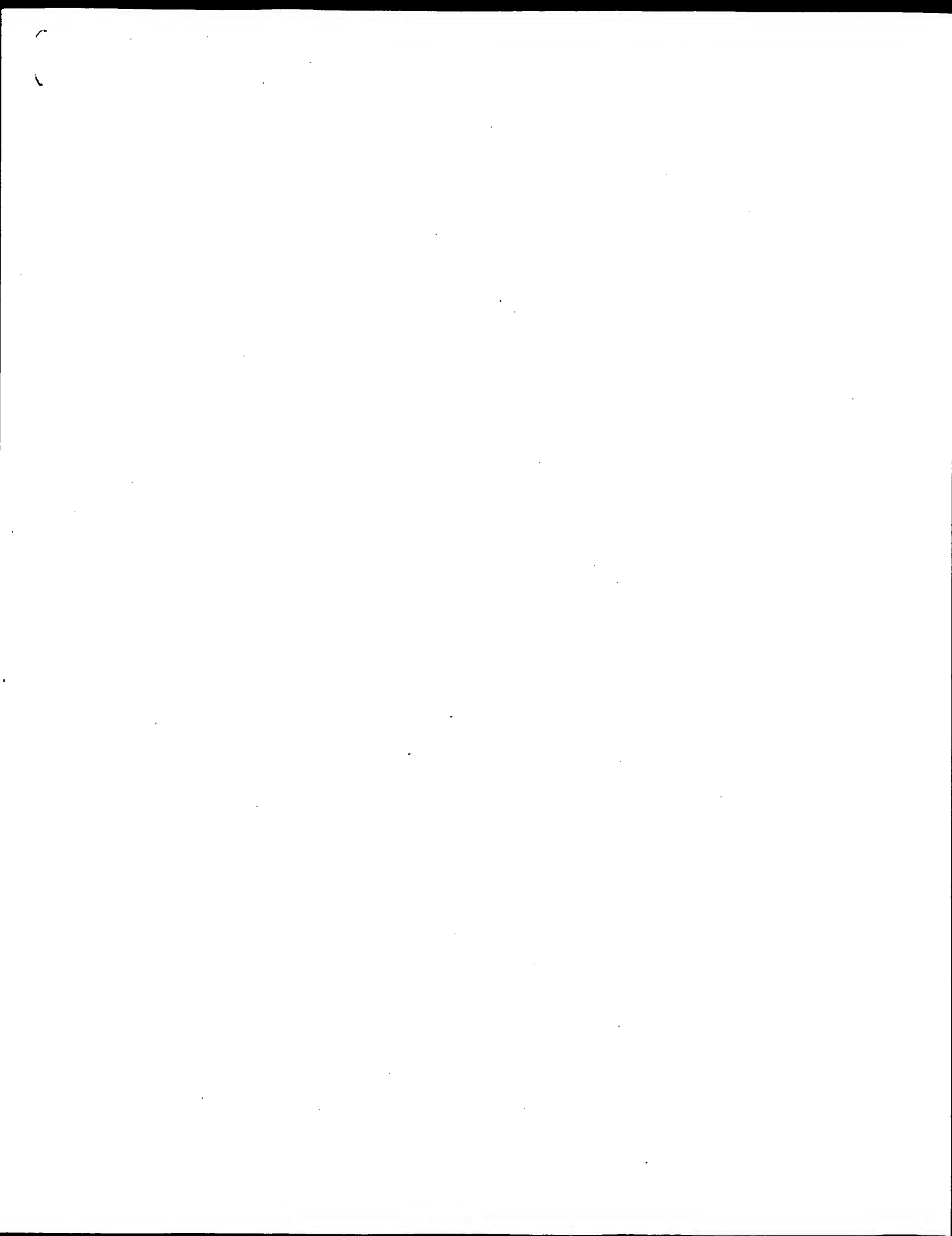
SPRMBL, Release 23.0, Released on 4Mar2003, Formatted on 7Mar2003

Total finds: 7

Total length: 305,079,309

Total sequences: 958,388

CPU time: 09:35.45



```

TOIG of: galt_cloab check: 8688 from: 1 to: 497
ID GALT_CLOAB STANDARD; PRT; 497 AA.
AC 097E24;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Galactose-1-phosphate uridylyltransferase (EC 2.7.7.12) (Gal-1-P
DE uridylyltransferase) (UDP-glucose--hexose-1-phosphate
DE uridylyltransferase).
GN GALT OR CAC2961.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RC MEDLINE=21359325; Pubmed=1146286;
RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RA "Genome sequence and comparative analysis of the solvent-producing
RA bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
CC -1- CATALYTIC ACTIVITY: UDP-glucose + alpha-D-galactose 1-phosphate =
CC -1- PATHWAY: galactose metabolism; second step.
CC -1- SUBCELLULAR LOCATION: cytoplasmic (Potential).
CC -1- SIMILARITY: Belongs to the galactose-1-phosphate
CC uridylyltransferase family 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE007793; AAK80903.1; -.
DR PIR: D97264; D97264.
DR HAMAP: MF_00571; -.
DR InterPro: IPR000766; GalP_transf_II.
DR InterPro: IPR005850; GalP_Utransf_C.
DR InterPro: IPR005849; GalP_Utransf_N.
DR InterPro: IPR005934; GalT2.
DR Pfam: PF02744; GalP_UDP_tr_C; 1.
DR Pfam: PF01087; GalP_UDP_transf_1.
DR TIGRfam: TIGR01239; galT2; 1.
DR PROSITE: PS01163; GAL_P_UDP_TRANSF_II; 1.
KW Transferase; Nucleotidyltransferase; Galactose metabolism;
KW complete proteome.
SQ SEQUENCE 497 AA; 57288 MW; 81E9D089CD9F5CC CRC64;
; GALT_CLOAB Length: 497 August 12, 2003 12:10 Type: P Check: 8688
; galt_cloab
MINHEIKKLLAFSLKGLIOEDDKYSSNMAGLNFNDYFEETSDVSTATAILNQLIAVAVKENLIN
DTVAERDLFDPTKIMCVPRSEYINNFNRLNNSPKKATSYKLSIASNYIKRDIRDNITWKPTPEY
GDLOITINLSPEEDPDHIAAKLSKSTSYKCLCKENEGFNGINHPAROTLRILPELNSKMFLOY
SPVTVYNEHCILNNEHILPMKISRTIFENLSFIDILPHFAGSNADLPVIGSILSHDYOGGRYFAM
EKAVEKEYSIKGYEDISVGVKWPMSVTRKTKLINLAELHLSWRNYSDDKOSTLSHSGSEPHN
TITFARRRNEEYELDILRNRRNTDENYPLGIFPHNVEHHIKKENIGLIEWGLAVLPARLKSSEALIK
ENLEKKKDSNDSTISKHNHWYIILDNKNISENDICLKKREVGKFLKLEVLKAGVFKRNSGSLAP
DKFINILL

```

```

TOIG of: 033907 check: 3609 from: 1 to: 543
ID 033907; PRELIMINARY; PRT; 543 AA.
AC 033907;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE Hypothetical 59.4 kDa protein.
GN Shewanella sp. SRC-2738.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=535560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SRC-2738.
RC MEDLINE=97419510; Pubmed=9274025;
RA Takeyama H., Takeda D., Yazawa K., Yamada A., Matsunaga T.;
RA "Expression of the eicosapentaenoic acid synthesis gene cluster from
RA Shewanella sp. in a transgenic marine cyanobacterium, Synechococcus
RA sp."
RL Microbiology 143:0-0(0).
DR EMBL: U73835; AAB81126.1; -.
KW Hypothetical protein.
SQ SEQUENCE 543 AA; 59378 MW; 7233F536358794C7 CRC64;
; 033907 Length: 543 August 12, 2003 12:10 Type: P Check: 3609
; 033907
MNPATFEMSLPMPNATFESNISFDQVMEQOLKDSRACVYVNHADHGFIAQTADIVTEQANSTDLR
VSAFTPALGTEISGDNNFRFVHGVKAYVYGAMANGISSEELVIALGOGAILCGSFGAAGLPSRVEAAI
NRIOALPNPMPYENLHSPSEPALERGSVELFLKHVRVESAFLGLTPQIVVYRAGLSHDAGKVV
VGNVIVAKVRTEVAEKFMAPAKMLQKLVDDGSITAEOMELAOVPMADDTIADDSGGHSDNRPIVY
LLEPIILAKKEIARAKYQYDPIIVGGGGVGTEDALATFNMGAAYIVGSIINAOVVEGASPHTRKLLA
TTAEADVTMAPADMEFMGKLVVVRGTFEPMRANKIYEIVYRVSIAIPIDDEKEXLEKOVFRSSLDE
IMGVYVHNERPQKQIERECPKRMALIFRWYIGLSRMSGSGVEGEMDYQIWAQPALGAFQMAK
GSIVTDNQDNNAVDLAHMLTGAYLRLINSLTAQGVKVPADLLRKKPMORMAL

```

```

TOIG of: p100_hsv6u check: 500 from: 1 to: 870
ID P100_HSV6U STANDARD: PRT: 870 AA.
AC 000701:
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Large structural phosphoprotein (Pp100) (P100) (Major antigenic
DE structural protein).
GN u11 OR P100.
OS Human herpesvirus (type 6 / strain Uganda-1102) (HHV6).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Roseolovirus.
OX NCBI_TaxID=10370;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92260671; PubMed=1374813;
RA Neipel F., Ellinger K., Pleckenstein B.;
RT "Gene for the major antigenic structural protein (p100) of human
RT herpesvirus 6."
RL J. Virol. 66:3918-3924(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94118404; PubMed=8289364;
RA Nicholas J., Martin M.;
RT "Nucleotide sequence analysis of a 38.5-kilobase-pair region of the
RT genome of human herpesvirus 6 encoding human cytomegalovirus";
RT Immediate-early gene homologs and transactivating functions."
RL J. Virol. 68:597-610(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95266321; PubMed=7747482;
RA Compes U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,
RA Martin M.E., Efstrathiou S., Craxton M., Macaulay H.A.;
RT "The DNA sequence of human herpesvirus-6: structure, coding content,
RT and genome evolution."
RL Virology 209:29-51(1995).
CC -1- SIMILARITY: TO THE LARGE STRUCTURAL PHOSPHOPROTEINS OF HSV-7 AND
CC HCV UL32.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: M87287; AAA46012.1; -
CC DR EMBL: L25528; AAA16716.1; ALT_INIT.
CC DR EMBL: X83413; CA58438.1; -
CC KW Matrix protein: Phosphorylation.
SQ SEQUENCE 870 AA; 97071 MW; F25954DEA19BF824 CRC64;
;
; P100_HSV6U Length: 870 August 12, 2003 12:10 Type: P Check: 500
;
; p100hsv6u
; MDLORHPDIPFAMLDKRDYERLTDLFLSLERLDVNDLREHPVNTNSCVREGGDVEDLKTLYNLLVLMY
; HYVALSKRPDYNAMODITLOSVNVEYNSKMGKIFENMFNKESQFSDIRALLRLGNFKMG
; SNVAIDPTPVNLTAEDESEIENNOADEKEMKMTWYVNIINDPDENGILYLSINKLILYGLKFLALQSM
; SKERVAMSQIVITONHLSGRLRHDFNIVYSHRVYQTLTGQVRESFKITISDYDIKSSLESAS
; KAFMSSEGPNSLMDPVLKRDHSNLTLPMSIDTKSSLDLPRLKRSRSLDFTLRQRPKLELD
; SYDNAGEKTLKEATLGEVNTKTPASSVSLMGVSSPSFTNLDPLSFTSTNLDLRKSHGNKY
; IGPGLIDENYKPPNQLQNTVADLQDTSITGSSGIDTVYNGFANLNLHQNSNVSPPWSRNTAAN
; ADPLDPRFVPEOTGPVYNNSDVAGSEAKHTITSTETGVSPRVFLIKDLRGKDGFRKOKSDIPKS
; LTRKRNDAKIHSEVTDGSDATETYGARNSPALRKIKQANFEAGLNKNDRDYLRGKGKSKDLSHG
; GNARKKESGKFNDDKENTRGQEPSPSKLMDNADBOYIQAGLQGRVNNLSQTNLISGEGKIED
; ILQNRGTLEKLATENKSGRESEANETILEVSNPDMEKFNRLQNDLDSVSPERLPADLSRELDISA
; SFKDALDLKPGNGEREIDLEKVKYGETETSDLKVGODESFVPAQLMKVETPEEKDDITIEQVNLIRO
; DGETDENVTSGPVAESIDLEAKGESAIASI

```

```

TOIG of: pfda_metka check: 9337 from: 1 to: 157
ID PFDA_METKA STANDARD: PRT: 157 AA.
AC 0810Y7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Prefoldin alpha subunit (Gimc alpha subunit).
DE PFDA OR MK1614.
GN Methanopyrus kandleri.
OS Archaea: Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OC NCBI_TaxID=2320;
OX [1]
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malyn A.G., Koonin E.V., Kozlovskiy S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
CC -1- FUNCTION: Molecular chaperone capable of stabilizing a range of
CC proteins. Seems to fulfil an ATP-independent, HSP70-like function
CC in archaeal de novo protein folding (By similarity).
CC -1- SUBUNIT: Heterohexamer of two alpha and four beta subunits (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PREFOLDIN ALPHA SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AE010451; AAM02827.1; -
CC DR HAMAP: MF_00308; -; 1.
CC DR InterPro: IPR004127; DUF232.
CC DR Pfam: PF02996; Prefoldin_1.
CC KW Chaperone; Complete proteome.
SQ SEQUENCE 157 AA; 17446 MW; B20BDA8CC978DFAL CRC64;
;
; PFDA_METKA Length: 157 August 12, 2003 12:10 Type: P Check: 9337
;
; pfda_metka
; pfdameetka
; MAEKNEQDITQGLRLIAEINRLQGMBAINAQDLESSTSELNRKYETLKVKEDEGEVIVPVA
; OSFVRACVDTERVIVGAGVAVERTIDEALESIDDORELERARABQKLOLRLAOLDEKORAKOEL
; AQLLEGAQRIFAQSSGSG1

```



```

TOIG of: q97wg2 check: 6793 from: 1 to: 280
ID 097wg2 PRELIMINARY; PRT: 220 AA.
AC 097wg2
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Hypothetical protein SSO2259.
GN SSO2259.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332286; PubMed=11427726;
RA She Q., Singh R.K., Contaloni F., Zivanovic Y., Allard G.,
RA Awayez M.J., Chan-Welner C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moers A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doollittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
RL "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
DR Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 220 AA; 24900 MW; EC65875E903DF2B CRC64;

```

```

097wg2 Length: 220 August 12, 2003 12:11 Type: P Check: 6793
MTNGEFYALDNLEKDEKLNKILVDINTTDRGLDVLKIGLEDENTIGKLTISLSDVLELVAMD
KVTKTKLFINDENYINQFLINLDKRSKGLIDPIIDKEDESLKAIINLNDPTLNNHWEIIT
NLSRIDLTNFEYITLVSATGEALKTENVRKTSIWEIYKLLKDPIDRGVAVASVAKIKIGLYVDK
GLAFVEKKLI
TOIG of: q9esp3 check: 4988 from: 1 to: 1851
ID 09ESP3 PRELIMINARY; PRT: 1851 AA.
AC 09ESP3
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 23, Last annotation update)
DE Mucin (Fragment).
GN MUC5AC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Stomach;
RA Ohnuma T., Suganuma T.;
RT "Rat gastric mucin Muc5AC: Sequence of its 5'-region contains
conserved D-domains and two leucine zipper motifs.";
DR EMBL, AB042530; BAB17787.1;
DR InterPro: IPR001064; Crystalin.
DR InterPro: IPR005829; Sug_transferrin.
DR InterPro: IPR002919; TIL_Cystich.
DR InterPro: IPR001846; VWF_D.
DR Pfam: PF01826; TIL; 3.
DR SMART: PF00094; vwd; 3.
DR SMART: SM00214; vwd; 3.
DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR PROSITE: PS00217; SUGAR_TRANSPORT_2; 1.
DR NON_TER 1851 1851
SQ SEQUENCE 1851 AA; 199403 MW; 879DE5B54929C52B CRC64;

```

```

09ESP3 Length: 1851 August 12, 2003 12:10 Type: P Check: 4988
q9esp3
MLHSMGVGRRLAPFVIALAFENQHTGOALEDTKRSIHEHYSDLSOPGHVGTPLNRYTIIPIKATP
VYRAFNPAHTRRVGCTGWNHFNHYKTFDGOVYFPGICGVYSEHGCAAYEDFNQLRGDESNTTISRYI
MKLDGLVVELTKSSVIVNNHPVOLPFSQSVLIELSGYLVARGLGVFMNDDSLLELDTKYANKT
CGLOGDNGSPESSEBELSHNVRLPLEFNGFQKMDGTEEOCDLPVPOKNCISIRSLCEILKQLFNS
CAALVDISSLYLEACQDCLCESDPSNCICHTLAESROCAHAGGQPMWRGNPLCPQCLIMTEYOEC
GSPCVDICSNPOHQQVEDHCYAGCEPEGMVLDSNQTGCVPSOACALYNGTLVAPGTSYDCTKCT
CSGGOMSCQOEYPCSGTSGVMGSHISPEDEROYTVHGDSCYLCRPYDSNAFTVLALERKCGLTSEEC
KTVTLNMGKGTVTYVKAAGEVFNQIYQTLQPVSTANAMPREPTFTIQTQNLGULEQLPPIQVSY
RIAEFRGLTSGLOGNNSQADDFQITISGVETGAAPFNTRITQACPMVANNIFEDPSLSVENEKVA
OHKCSQITDANGPPSOCHAVNPSSTFSCNMFDTNCNEKSEDELCALASVYVACAKAVGLLSDMREGIC
AKPTITCPKSMTYOYHISTQPCRSLSBEDVTCNFIPEVDCGTPKGFSLDSCGVQATSCPCYRG
SPVNGESVHNDNGAICCTOGALTCIGGPAVLFEVCDAPMIYFDCRNATPGDTAGCOKSCHTIDMTCVSS
ECVPGCYVNGLVADNGSCVVAEDPCVNEATYRGETIOVCNCTCENRMMQCTDKPCLATCAVYG
DGHYITFDGORYSFGSCETYLIDONGNGSSODAFRVYTEMIPGCTGCTGCTGCTGCTGCTGCTGCTG
SKMEVVGKGOEYRPGVHOMGNVLYEMDGLVILMDKTSIFLALSPERFKVAGGLCGNEDFNKINDE
TTRSOVSMDLEFGNEMKLSPCSPDASVSKDCTAPTRKSNAGQCSITINSAAVSAACHAHEPAKYE
ACVNDACACDSGGGCECFCTAVAAVAAQACHEVGVCSWTRPDICPLECDYNGEGCEMHWYOCGACMK
TCQPTGQLODLRGLEGECYKPCPTAPIDEQTMOCVSNCTVSPCERNGLYAPGPISDSNCTYC
VCTESGVNCTHDAGACACVCTYNGRIGDITTYTDMGGCISAHORNGTIRIYDTSSTSPPTTF
SFTTILVMTMOPSSSTHSPPTSVVYPGSPSKAVLASSVSVKPTETSVLTTSASATIMPAOEC
LMSFMDISRGRIGDGDPTLENLAHGQYOCVPAKAYECRADENPCVAFHALOHNCESTYVILICY
NSDOVSGICDNVOIKIOCCPCINCPSTGPTOHLIVASTSMEETSSVPTSTERYHVIASVSPTH
TPGSPSSSVYSSSAPARSTPTVYSTTAVKTLPTTSPEPEPSSVSVSISTLSTLASPEYTHCKRE
LCNMNDLPIPCVYENIRIECCITLVNICSTSATQPTSGVSIKTKTNNTNTYTSSTENTSGHSTVINT
KTVVTEGTHHTTPOGTRPPTSVISVOTDTSSTSSHTSSPNTGRVSTHTTHTSSPTGCTSI
TSTHTSSPNTGCTSIPTHTHTSSPTGCTSI

```

TOIG of: yzn4\_caee1 check: 9797 from: 1 to: 692

ID yzn4\_caee1 STANDARD: PRT: 692 AA.

AC P54245;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical acetylcholine receptor like protein F18G5.4 in

DE Chromosome X.

GN F18G5.4.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-Bristol N2;

RA Favella A.;

RL Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.

CC -1- FUNCTION: POSSIBLE ACETYLCHOLINE RECEPTOR.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

DR EMBL: U98855; AAA81080.1; .

DR PIR: A89606; A89606.

DR WormPep: F18G5.4; CE04412.

DR InterPro: IPR006029; Neur\_channel\_memb.

DR InterPro: IPR006202; Neur\_chan\_LBD.

DR InterPro: IPR006201; Neur\_channel.

DR Pfam: PF02931; Neur\_chan\_LBD; 1.

DR Pfam: PF02932; Neur\_chan\_memb; 1.

DR PROSITE: PS00236; NEURORF\_ION\_CHANNEL; 1.

KW Hypothetical protein; Receptor; Postsynaptic membrane; Ionic channel;

KW Glycoprotein; Transmembrane.

FT TRANSMEM 39 59 POTENTIAL.

FT TRANSMEM 402 422 POTENTIAL.

FT TRANSMEM 432 452 POTENTIAL.

FT TRANSMEM 460 481 POTENTIAL.

FT DISULFID 315 329 BY SIMILARITY.

FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 593 593 N-LINKED (GLCNAC. . .) (POTENTIAL).

SO SEQUENCE 692 AA: 79291 MW: 3510664780FPAD27 CRG64;

.. yzn4\_caee1 length: 692 August 12, 2003 12:10 Type: P Check: 9797 ..

.. yzn4\_caee1

MKVASLAFHCFHGLRCLKPRHKLPEPRHSKLVRALLPLLSHPAFEFLECAPFHTITFALHITY

VRRREDLLILLPLGTCOPAHHPRSFNSNMLLKROSVOSTGPOPPKPKOPKPHNEENTIGITKRAPSVQEOHS

SAVIMPFDFONRLQALRIGSDGTREALYRSLDHTYERKIDYVRCIHHSOPTNWTFGFLNOIYEMD

ERNQALITRSMNLINWMDPRISWNSSEITKAITIPARIWKPDIILVNNKOTKNSDACIHAITISEP

AAIRRYVASLSTDMVTSDGNVTLFSAFLRSSCPIRVRYPPDDQCDLKFASWSDITELNLGLNTD

KGLDSSYNNSEFIDLVDMTAVREYVTPSPDINSMDPIIVIRIHMHRPLFVFNHIVPCVLIISMAVIGF

LMPEPTGEKINMLITTLISMGVLOSITESIPTSEGVGLIYGYVSSLVWCLATCVNLTILNMHRNGA

ANOGHVAFAMMOKOTIGLATFEMMSIREPDSIALKASOSKSTIRRSSILRDLKRYKNNSVNRASKE

ONANRECCMPPIVHIYVATMSCLADPTKPMNSSTIREDPASSTPLGRVSGIMPRISASNSVYTE

FETRFRILKRYNSLQOHEITREITLDRSRIQCSGNMLSLIDFTYVFPALQHCSSSASI